ISSN : 2321-9602

Indo-American Journal of Agricultural and Veterinary Sciences




editor@iajavs.com
iajavs.editor@gmail.com

Indo-American Journal of Agricultural and Veterinary Sciences

Vol. 10, Issue 3 July 2022

NETWORK TRAFFIC ANALYSIS USING MACHINE LEARNING

BONAGIRI AKHILA1, DULAM SWATHI2, CHILUMULA KARTHIK3, MARSHETTY SAITEJA4

ABSTRACT:
In the field of networking, it is often necessary to identify the sorts of applications that traverse a network in order to complete certain tasks. Classification of network traffic is primarily used by Internet service providers (ISPs) to examine the characteristics necessary for network architecture, which impacts the network's overall performance. There are several classification methods for network protocols, such as port-based, payload-based, and Machine Learning-based, and each has its own advantages and disadvantages. Due to its widespread use in different domains and academics' rising awareness of its superior accuracy when compared to others, the Machine Learning approach is now prominent. In this research, we examine the performance of two fundamental algorithms, Nave Bayes and K nearest, when applied to networking data retrieved from live video stream using the Wireshark program. Python's sklearn library is used to implement the Machine Learning algorithm, together with the numpy and pandas libraries as assistance libraries. Finally, we see that the K closest approach provides more precise predictions than the Naive Bayes, Decision Tree, and Support Vector Machine algorithms.

Keywords: Bayes, wireshark, and the K closest algorithm.

1. INTRODUCTION:

As a result of the proliferation of smart devices, networks have grown more diverse and dynamic, compelling network operators to seek for new management methods. Therefore, developing network architecture capable of handling the maximizing heterogeneity and resource use efficiency is a significant problem [1]. Several solutions have been offered in this area, including Network Slicing (NS) and Machine Learning (ML). Together, these approaches may enable intelligent and autonomous network resource management, as well as performance gains. [2] The optimization of a large-scale environment to meet 5G quality of service (QoS) criteria. In reality, ML is used to handle difficult issues without explicit programming, where the algorithm may model and understand the underlying behavior by using a training dataset/environment. Its efficacy has been shown and it has yielded encouraging outcomes in other fields, including network management [3]. NS, on the other hand, offers network as a service (NaaS) for many use cases, enabling infrastructure providers to make their physical network accessible to multiple tenants.

1,2, 3, 4 UG Student, Department of CSE, NARSIMHA REDDY ENGINEERING COLLEGE, Maisammaguda, Kompally, Secunderabad, Telangana India. 500100

[4] [5]. It divides the network infrastructure into network slices, each with its own resources and performance requirements. NS provides several benefits for 5G networks and beyond. First, it permits several tenants to use the same physical network infrastructure. Second, NS is capable of managing diverse services with rigorous QoS and service level agreements (SLA). Thirdly, since it handles diverse service needs as opposed to the "one-size-fits-all" principle [4, it results in a flexible and efficient use of limited resources].

Nevertheless, defining the network state explicitly is a difficult process due to the fact that network behavior varies based on characteristics such as user mobility, geography, and social events. In this situation, allocating resources to network slicing instances becomes more challenging. In addition, the enormous number of new network traffic and apps made the manual generation of network slices cumbersome [6]. Therefore, the process should be adaptable, responsive to instantaneous demands, and dependent on the end-traffic user's pattern. Using intelligent traffic management, which comprehends the behavior of linked smart devices and apps, it is possible to monitor the performance of network slices and optimize network resources (e.g., auto-scaling). Therefore, the intelligent design of Network (sub)-Slicing based on traffic patterns will become a crucial problem for supporting diverse use cases in varied settings and even inside the slice itself (sub-slicing).

ML provides these advantages for quickly understanding traffic and automatically generating network slicing since it can examine a big amount of data and identify a suitable pattern.

This data can be derived in a fair amount of time. It enables the system to generate rules to automate processes. The most significant benefit of ML is its capacity to solve complicated problems. In addition, the study of these massive datasets using machine learning offers fresh findings. Consequently, incorporating these technologies into traffic analysis and NS will allow network operators to deploy self-configuring, self-healing, and self-optimizing networks [7], so enabling zero-touch administration. In the realm of machine learning, there is a clear contrast between supervised and unsupervised learning. The primary goal of supervised learning is to establish a mapping between the input characteristics and the output class, which needs a labeled data set. The purpose of unsupervised learning, on the other hand, is to detect a structure (pattern) in the inputs without the necessity for an output class.

In reality, machine learning models can only learn as well as the data and characteristics provided. The increase in data dimensionality may reduce the efficiency of an algorithm and incur additional computing costs (such as storage and processing) [8]. and bane of dimensionality issue In order to make the raw data appropriate for analysis, preprocessing techniques must be undertaken, one of which is feature selection.

In reality, machine learning models can only learn as well as the data and characteristics provided. The increase in data dimensionality may reduce the efficiency of an algorithm and incur more computing costs (such as storage and processing) [8] and the curse of dimensionality issue. In order to make the raw data appropriate for analysis, preprocessing techniques must be undertaken, one of which is feature selection.

LITERATURE SURVEY

Next-generation wireless networks using AI-assisted network slicing: Next-generation wireless networks (NGWNs) are very heterogeneous and dynamic due to the integration of communications with varying sizes, multiple radio access methods, and numerous network resources. Emerging use cases and applications, including as machine-to-machine communications, autonomous driving, and industrial automation, have strict dependability, latency, and throughput requirements. throughput, among etc. Consequently, architectural design, network administration, and resource orchestration in NGWNs face significant issues. Beginning with an illustration of these obstacles, this study intends to provide a comprehensive overview of the general design of NGWNs as well as three particular research topics pertaining to this architecture. Then, we describe why and where artificial intelligence (AI) should be implemented into this network-slicing-based design. Second, the motivation, research challenges, existing works, and potential future directions associated with the application of AI-based approaches to three research problems, namely flexible radio access network slicing, automated radio access technology selection, and mobile edge caching and content delivery, are described in detail. This article summarizes the merits and potentials of AI-based techniques in the study of NGWNs.

When network slicing meets the notion of prospects: A framework for service provider revenue maximization:

Recently, network slicing technology has evolved, allowing network operators to provide specialized virtual networks. clients, over a shared network infrastructure. This study examines the subject of service provider revenue maximization in network slicing, with the goal of providing tailored services for various user classes and diverse needs. Specifically, the network is sliced based on the viewpoint of the end user, which is examined at many levels. Consumers' demand forecasting for service classes is pursued via the implementation of a federated learning architecture in which end users serve as customers. Then, taking into account the anticipated service needs, the virtual network functions placement strategy is executed by using matching theory principles while taking into mind that various network zones are defined by varying provision costs and prices. In order to create a realistic users' decision-making procedure, the ideas of prospect theory are also utilized in order to account for the customers' views. Last but not least, the validity of the proposed framework is confirmed by presenting numerical findings generated from extensive computer simulations. comparing the performance of the Kolkata game with the hypothetical game.

EXISTING SYSTEM:

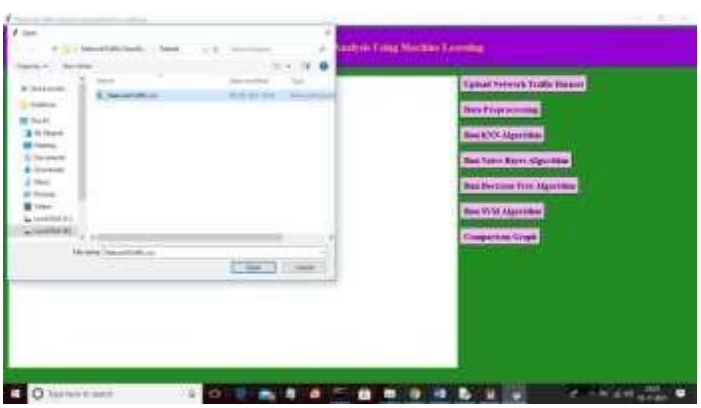
In the field of networking, it is often necessary to identify the sorts of applications that traverse a network in order to complete certain tasks. Classification of network traffic is primarily used by Internet service providers (ISPs) to examine the characteristics necessary for network architecture, which impacts the network's overall performance. There are several ways used to categorize network protocols, such as port-based, payload-based, and Machine Learning-based classification, and each has its own advantages and disadvantages. Due to its widespread use in different domains and academics' rising awareness of its superior accuracy when compared to others, the Machine Learning approach is now prominent.

PROPOSED SYSTEM:

We examine the performance of two fundamental algorithms, Nave Bayes and K closest, when applied to networking data retrieved from a live video stream using the Wireshark program. Python is used to implement the Machine Learning algorithm.

In conjunction with numpy and pandas libraries, the sklearn library is used. Finally, we see that the K closest approach provides more precise predictions than the Naive Bayes, Decision Tree, and Support Vector Machine algorithms.

To execute the project, double-click the "run.bat" file.



Click the "upload Network Traffic Dataset" button on the aforementioned screen to upload the dataset.

In the previous page, choose and upload the "NetworkTraffic.csv" file, then click the "Open" button to load the dataset and access the subsequent screen. In the previous page, choose and upload the "NetworkTraffic.csv" file, then click the "Open" button to load the

dataset and access the subsequent screen.

This dataset has many non-numeric variables, thus it must be processed. The x-axis of the graph shows the kind of traffic, while the y-axis reflects the number of entries in the dataset for that type of traffic. Now exit the preceding graph and click "Data Preprocessing" to clean the dataset.

On the screen seen above, the whole dataset has been transformed to numeric format, after which the total number of records and columns present in the dataset, as well as the proportion of train and test records, are displayed. Now that train and test data are prepared, click the "Run KNN Algorithm" button to train KNN and get the results shown below.

Ir
n
a
ir

CONCLUSION

In order to define network slicing, this research has been conducted as a proof of concept integrating ML with traffic analysis. This research shows how machine learning has the potential to design intelligent network slices by analyzing traffic behaviour. This uses the Feature Selection (RFE) technique and the K-means algorithm. Proposal to determine the most pertinent clustering and features, respectively. To readily be used for network slicing and resource management, an experimental investigation of these clusters (future slices) has been undertaken using the chosen attributes to discover relevant behaviors. These properties are divided into categories based on their application-, time-, and bandwidth-related characteristics. The data collected have clearly shown several habits that need particular attention. For instance, in the

morning of 04/27, clusters 0 and 1 (future slices) need greater resources from the infrastructure provider due to the increased traffic demand at this time. Additionally, as proxy traffic makes up the bulk of cluster 0's traffic, extra attention should be paid to resource allocation for this cluster (for example, providing the proxy server with greater CPU and cache resources) (future slice). As a result, this enables the infrastructure provider to foresee network status and prevent SLA violations. In order to compare the performance to the K-mean clustering employed in this article, future work should also use additional clustering methods, such as DBSCAN (Densitybased spatial clustering of applications with noise) or hierarchical clustering, to the dataset. It is also important to implement network slicing infrastructure in order to benchmark the signaling and overhead produced by this method.

REFERANCES

1. "Aias aided network- slicing based next-generation wireless networks," IEEE Open Journal of Vehicular Technology, vol. 1, pp. 45-66, 2020. X. Shen, J. Gao, W. Wu, K. Lyu, M. Li, W. Zhuang, X. Li, and J. Rao

2. "When network slicing meets prospect theory: A service provider revenue maximization paradigm," R. Fantacci and B. Picano, IEEE Transactions on Vehicular Technology, vol. 69, no. 3, 2020, pp. 3179-3189

3. F. Estrada-Solano, N. Limam, S. Ayoubi, N. Shahriar, R. Boutaba, M. A. Salahuddin, and O.Journal of Internet Services and Applications, vol. 9, no. 1, pp. 1-99, M. Caicedo, "A Comprehensive Survey on Machine Learning for Networking: Evolution, Applications, and Research Opportunities," 2018.

4. X. Li, D. Bhamare, M. Samaka, H. A. Chan Network slicing for 5G: Challenges and Opportunities, L. Gupta, C. Guo, and R. Jain, IEEE Internet Computing, vol. 21, no. 5, pp. 20–27, 2017.

5. "Optimal 5G network slicing utilizing machine learning and deep learning ideas," Computer Standards & Interfaces, p. 103518, 2021. M. H. Abidi, H. Alkhalefah, K. Moiduddin, M. Alazab, M. K. Mohammed, W. Ameen, and T. R. Gadekallu

6. "Consideration on automation of 5G network slicing using machine learning," in 2018 ITU Kaleidoscope: Machine Learning for a 5G Future (ITU K), pp. 1-8, IEEE, by V. P. Kafle, Y. Fukushima, P. Martinez-Julia, and T. Miyazawa.

7. "Knowledge-defined networking," ACM SIGCOMM Computer Communication Review, vol. 47, no. 3, pp. 2-10, 2017. A. Mestres, A. Rodriguez-Natal, J. Carner, P. Barlet-Ros, E. Alarc'on, M. Sol'e, V. Munt'es-Mulero, D. Meyer, S. Barkai, M. J. Hibbett, et al.

8. "Machine learning with large data: Challenges and techniques," Ieee Access, vol. 5, pp. 7776-7797, 2017. A. Lheureux, K. Grolinger, H. F. Elyamany, and M. A. Capretz

9. "Network traffic categorization using machine learning for software defined networks," in International Conference on Machine Learning for Networking, Springer, 2019, pp. 28–39. M. P. J. Kuranage, K. Piamrat, and S. Hamma.

10. "Sdn/nfv, machine learning, and big data driven network slicing for 5g," in 2018 IEEE 5G World Forum (5GWF), pp. 20–25, IEEE, by L.-V. Le, B.-S. P. Lin, L.-P. Tung, and D. Sinh.

11. "Toward in-network deep machine learning for recognizing mobile apps and allowing application specific network slicing" by A. Nakao and P. Du, IEICE Transactions on Communications, pp. 1536-1543, 2018.